

MSW Spaces*

H. A. MAURER

*Institut für Informationsverarbeitung, Technische Universität Graz, Steyrergasse 17,
A-8010 Graz, Austria*

A. SALOMAA

Department of Mathematics, University of Turku, SF-20500 Turku 50, Finland

AND

D. WOOD

Unit for Computer Science, McMaster University, Hamilton, Ontario L8S 4K1, Canada

This paper is a first attempt to systematically study collections of language families which are *dense* in the sense that between any two families one can "squeeze in" another one. By considering collections of language families satisfying some mild restrictions we obtain the notion of an MSW space. We show that many such spaces exist and that each MSW space is dense "above" a sufficiently large language family.

1. INTRODUCTION

In language theory, the relative position of language classes has been one of the major concerns. Although many results concerning individual language families have been obtained (such as one specific family being properly contained in some other specific family) very little is known about (infinite) collections of language families. Even concerning some much investigated collections of language families such as \mathcal{M}_{AFL} , the collection of all AFLs, or \mathcal{M}_{cone} , the collection of all rational cones, many of the most basic results remain open. For instance, it is still not known whether there is a smallest AFL properly containing the regular languages.

In form theory, cf. Wood (1979) for a recent survey, collections of language arise naturally. Recently, collections \mathcal{M} have been found in Maurer *et al.*

* Research in this paper was supported by a Natural Sciences and Engineering Research Council of Canada Grant A-7700 and by the Austrian Federal Ministry of Science and Research.

(1979c, d) which are *dense* in the sense that between any two language families \mathcal{L}_1 and \mathcal{L}_2 of \mathcal{M} with $\mathcal{L}_1 \subsetneq \mathcal{L}_2$ a language family \mathcal{L}_3 of \mathcal{M} with $\mathcal{L}_1 \subsetneq \mathcal{L}_3 \subsetneq \mathcal{L}_2$ can be found. The first instance of such a collection discovered in language theory is the collection

$$\mathcal{M} = \{\mathcal{L}(F) : F \text{ is a context free grammar form and } \mathcal{L}(F) \text{ contains all finite languages}\}$$

discussed in detail in Maurer *et al.* (1979c).

Encouraged by this result, we investigate in this paper collections of language families giving rise to such density properties.

We introduce an MSW space as a collection of language families satisfying some mild assumptions. We then show that in any MSW space \mathcal{M} the collection of language families $\{\mathcal{L} \in \mathcal{M} : \mathcal{L}_0 \subseteq \mathcal{L}\}$ is dense, provided \mathcal{L}_0 is large enough.

We derive a number of results which show that there is an abundance of MSW spaces. Hence, the phenomenon of density first defined by generative devices is not specific to that theory but a natural property of many collections of language families.

2. PRELIMINARIES

In this section we introduce a number of notions required in the sequel, many of them already discussed in Maurer *et al.* (1979c). We also briefly review some well-known notations and definitions from formal language theory but refer to either of Harrison (1978), Maurer (1969) or Salomaa (1975) for further details. In some of the examples (but only in the examples) definitions and results from form theory are used. For completeness sake we explain the notion of a context-free grammar form and its interpretations (in the sense of Maurer *et al.* (1980a)) as used in Example 1 of Section 3. For non-context-free grammar forms we refer to Maurer *et al.* (1979b) and Maurer *et al.* (1979), for grammar forms in the sense of Cremers and Ginsburg to Cremers and Ginsburg (1975), for EOL forms to Maurer *et al.* (1977) and for AFL theory to Ginsburg (1975).

We start by discussing a number of operations on languages and language families.

Let L_1 and L_2 be languages over disjoint alphabets. Their *superdisjoint union*, denoted by $L_1 \dot{\cup} L_2$, is just the union of L_1 and L_2 . Observe that the terminology "superdisjoint union" and the notation $\dot{\cup}$ serve to specify that the operation is defined only if the alphabets of the languages involved are disjoint.

We now define the operation of breaking as a kind of inverse of the operation superdisjoint union. Let L be a language over some alphabet Σ . The language L_1 is obtained from L by *breaking* (with respect to an alphabet $\Sigma_1 \subseteq \Sigma$) if $L_1 = L \cap \Sigma_1^*$ and $L - L_1$ contains no word containing a symbol of Σ_1 . A language L

is called *coherent* if it cannot be broken in a nontrivial fashion; more precisely, if L_1 is obtained from L by breaking then either $L_1 = L$ or $L_1 = \emptyset$ (the empty set).

Throughout this paper, for every language L and integer $i \geq 1$ we denote by $L(i)$ the language defined by $L(i) = \{x \in L : |x| \neq i\}$. That is, $L(i)$ consists of all words of L whose length is different from i . Similarly, for an arbitrary language family \mathcal{L} and integer $i \geq 1$ we denote by $\mathcal{L}(i)$ the language family $\mathcal{L}(i) = \{L(i) : L \in \mathcal{L}\}$ and we call $\mathcal{L}(i)$ an *extraction* of \mathcal{L} .

A language family \mathcal{L} is closed under *covering* if for every infinite language L the fact that $L(i)$ is in \mathcal{L} for infinitely many i implies that L itself is also in \mathcal{L} .

The *superdisjoint wedge* of two language families \mathcal{L}_1 and \mathcal{L}_2 , in symbols $\mathcal{L}_1 \dot{\vee} \mathcal{L}_2$, is defined by $\mathcal{L}_1 \dot{\vee} \mathcal{L}_2 = \{L_1 \dot{\cup} L_2 : L_1 \in \mathcal{L}_1, L_2 \in \mathcal{L}_2\}$.

We now formulate the notion of dense collections of language families as first considered in Maurer *et al.* (1979c) and Maurer *et al.* (1979d).

Let \mathcal{M} be a collection of language families. \mathcal{M} is called *dense* if for any two language families \mathcal{L}_1 and \mathcal{L}_2 in \mathcal{M} with $\mathcal{L}_1 \subsetneq \mathcal{L}_2$ there exist a language family $\mathcal{L}_3 \in \mathcal{M}$ strictly in between, i.e., $\mathcal{L}_1 \subsetneq \mathcal{L}_3 \subsetneq \mathcal{L}_2$. Two language families $\mathcal{L}_1, \mathcal{L}_2$ of \mathcal{M} with $\mathcal{L}_1 \subsetneq \mathcal{L}_2$ are called a *dense pair* (with respect to \mathcal{M}) if $\{\mathcal{L} \in \mathcal{M} : \mathcal{L}_1 \subseteq \mathcal{L} \subseteq \mathcal{L}_2\}$ is dense. \mathcal{L}_1 is called *density forcing* (with respect to \mathcal{M}) if $\{\mathcal{L} \in \mathcal{M} : \mathcal{L}_1 \subseteq \mathcal{L}\}$ is dense.

For some examples in the next section we need notions in connection with grammar forms, cf. Maurer *et al.* (1980a).

Let V and V' be alphabets. A substitution μ defined on V is called a *df-substitution* if $\mu(a) \subseteq V'$ for every $a \in V$ and $\mu(a) \cap \mu(b) = \emptyset$ for all $a, b \in V$ with $a \neq b$.

Context-free grammars, CF grammars for short, are defined in the usual manner as quadruples $G = (V, \Sigma, P, S)$, where V is the total alphabet, $\Sigma \subseteq V$ the set of terminals, P the set of productions and $S \in V - \Sigma$ the start symbol. The language $L(G)$ generated by G is defined as usual. A *CF-grammar form* F is just a CF grammar. Let $F = (V, \Sigma, P, S)$ and $F' = (V', \Sigma', P', S')$ be CF grammar forms, and let μ be a df-substitution. F' is an *interpretation* of F (modulo μ), in symbols $F' \triangleleft F(\mu)$ (or just $F' \triangleleft F$ if μ is understood) if the following conditions (i)–(iii) hold:

- (i) $S' \in \mu(S)$,
- (ii) $\Sigma' \subseteq \mu(\Sigma)$, $V' - \Sigma' \subseteq \mu(V - \Sigma)$,
- (iii) $P' \subseteq \mu(P)$, where $\mu(P) = \{B \rightarrow y : A \rightarrow x \in P, B \in \mu(A), y \in \mu(x)\}$.

Let F be a CF grammar form. The *family of grammars generated by F* , denoted by $\mathcal{G}(F)$, is defined by $\mathcal{G}(F) = \{F' : F' \triangleleft F\}$. The *family of languages generated by F* , denoted by $\mathcal{L}(F)$, is defined by $\mathcal{L}(F) = \{L(F') : F' \triangleleft F\}$.

Other types of interpretation mechanisms have been considered in the past, in particular in Cremers and Ginsburg (1975) and follow-up papers thereof.

The notion of *form* can also be extended in the obvious way to the non-context-free case, see, e.g., Maurer, *et al.* (1979), Maurer *et al.* (1979b), to EOL-systems, see, e.g., Maurer *et al.* (1977) and to others. For further details and references consult Wood (1979) or Wood (1980).

As is often done, \mathcal{L}_{Fin} , \mathcal{L}_{Reg} , \mathcal{L}_{Lin} , \mathcal{L}_{CF} will denote the families of finite, regular, linear and context-free languages, respectively. Finally, we define one operation on CF grammars.

Let $F_i = (V_i, \Sigma_i, P_i, S_i)$ be CF grammars such that S_i does not occur on the right-hand side of any production of P_i ($i = 1, 2$) and suppose $V_1 \cap V_2 = \emptyset$. Define a new CF grammar $F_1 \oplus F_2 = (\bar{V}, \bar{\Sigma}, \bar{P}, S_1)$, where $\bar{V} = V_1 \cup V_2 - \{S_2\}$, $\bar{\Sigma} = \Sigma_1 \cup \Sigma_2$, $\bar{P} = P_1 \cup P_2$, where P_2' is P_2 with the nonterminal S_2 replaced by S_1 . Observe that $L(F_1 \oplus F_2) = L(F_1) \dot{\cup} L(F_2)$ and that $F_1 \triangleleft F$, $F_2 \triangleleft F$ implies $F_1 \oplus F_2 \triangleleft F$.

3. MSW SPACES

We start by defining the notion of an MSW space and by presenting a number of examples of such spaces. We then prove (Theorem 3.1) that in an MSW space every language family which is "large enough" is density forcing. We next present a result (Theorem 3.2) which establishes the existence of a variety of MSW spaces. After demonstrating a number of applications we present a theorem (Theorem 3.10) which makes the construction of MSW spaces particularly easy.

DEFINITION. A collection \mathcal{M} of language families is an *MSW space* if the following conditions (i)–(iii) hold:

- (i) Each \mathcal{L} in \mathcal{M} is closed under superdisjoint union and breaking.
- (ii) \mathcal{M} is closed under superdisjoint wedge.
- (iii) For each infinite language L occurring in some language family of \mathcal{M} there exist subsets L_i of L for $i = 1, 2, \dots$ such that each of $\{L_i : 1 \leq i \leq n\}$ is infinite and (a) and (b) hold:
 - (a) L is in a language family \mathcal{L} of \mathcal{M} iff L_i is in \mathcal{L} for all i with $L_i \neq L$.
 - (b) If L belongs to $\mathcal{L} \in \mathcal{M}$, then for every p with $L_p \neq L$ there exists an $\mathcal{L}_p \in \mathcal{M}$ such that $\mathcal{L}_p \subseteq \mathcal{L}$ and \mathcal{L}_p contains L_p but does not contain L .

We now discuss a number of examples of MSW spaces.

EXAMPLE 1. The collection of language families defined by $\mathcal{M} = \{\mathcal{L}(F) : F \text{ is a CF grammar form}\}$ is an MSW space.

Proof. (i) Consider an arbitrary $\mathcal{L} \in \mathcal{M}$. We may assume that $\mathcal{L} = \mathcal{L}(F)$, where $F = (V, \Sigma, P, S)$ is a CF grammar form in which S does not occur on

the right hand side of any production. Consider two arbitrary languages L_1, L_2 of \mathcal{L} over disjoint alphabets Σ_1, Σ_2 . We may assume that $L_i = L(F_i)$, where $F_i = (V_i, \Sigma_i, P_i, S_i) \triangleleft F$ for $i = 1, 2$ and where $V_1 \cap V_2 = \emptyset$ holds. Consider now the CF grammar form $F_1 \oplus F_2$ as defined at the end of Section 2. Since $F_1 \oplus F_2 \triangleleft F$ and $L(F_1 \oplus F_2) = L(F_1) \dot{\cup} L(F_2)$, \mathcal{L} is closed under superdisjoint union. That \mathcal{L} is closed under breaking follows from the fact that $\mathcal{L} = \mathcal{L}(F)$ is closed under intersection with regular sets for every CF grammar form F , cf. Maurer *et al.* (1980a).

(ii) Consider two language families $\mathcal{L}_1, \mathcal{L}_2$ in \mathcal{M} . We may assume $\mathcal{L}_i = \mathcal{L}(F_i)$, where $F_i = (V_i, \Sigma_i, P_i, S_i)$ are CF grammar forms for $i = 1, 2$ such that $V_1 \cap V_2 = \emptyset$. Consider the grammar form $F = (V_1 \cup V_2 \cup \{S\}, \Sigma_1 \cup \Sigma_2, P, S)$, where S is a new nonterminal, and where $P = P_1 \cup P_2 \cup \{S \rightarrow S_1, S \rightarrow S_2\}$. Evidently, $\mathcal{L}(F) = \{L_1 \dot{\cup} L_2 : L_1 \in \mathcal{L}_1, L_2 \in \mathcal{L}_2\} = \mathcal{L}_1 \dot{\vee} \mathcal{L}_2$. That is, \mathcal{M} is closed under superdisjoint wedge.

(iii) For each language L and each $i \geq 1$ define $L_i = L(i) = \{x \in L : |x| \neq i\}$. We will show that (a) and (b) are satisfied.

(a) Suppose $L \in \mathcal{L} = \mathcal{L}(F)$, F a CF grammar form. Since $\mathcal{L}(F)$ is known to be closed under intersection by regular sets, each $L_i \in \mathcal{L}$ for $i \geq 1$. Suppose all $L_i \neq L$ are in \mathcal{L} . Then $L_i = L(F_i)$, $F_i \triangleleft F(\mu_i)$ holds for infinitely many i , i.e., $L_i \subseteq \mu_i(L(F))$ for infinitely many i . Since the terminal alphabets of $L(F)$ and of L are fixed, for some i, j with $i \neq j$ we have $\mu_i = \mu_j$ as far as terminals are concerned. Thus $F_i \triangleleft F(\mu_i)$ and $F_j \triangleleft F(\mu_i)$. Hence we can clearly construct an $\bar{F} \triangleleft F$ such that $L(\bar{F}) = L_i \cup L_j$. But $L_i \cup L_j = L$ and we have obtained $L \in \mathcal{L}(F)$ as desired.

(b) Consider an arbitrary $L \in \mathcal{L} \in \mathcal{M}$ and an arbitrary $p \geq 1$ such that $L_p \neq L$. Consider the language family \mathcal{L}_p defined by $\mathcal{L}_p = \mathcal{L}(p) = \{L(p) : L \in \mathcal{L}\}$.

Since \mathcal{L} is closed under intersection by regular sets, $\mathcal{L}_p \subseteq \mathcal{L}$. Indeed, from F with $\mathcal{L}(F) = \mathcal{L}$ we can determine an interpretation $F_p \triangleleft F$ such that $\mathcal{L}_p = \mathcal{L}(F_p)$, i.e., $\mathcal{L}_p \in \mathcal{M}$. By construction, $L \in \mathcal{L}$ and hence $L_p \in \mathcal{L}_p$. Since $L_p \neq L$, L must contain a word of length p , but no language of L_p contains a word of length p , i.e., $L \notin \mathcal{L}_p$. This concludes the proof that \mathcal{M} is an MSW space. ■

EXAMPLE 2. Each of the following collections $\mathcal{M}_1 - \mathcal{M}_5$ are also MSW spaces.

- (1) $\mathcal{M}_1 = \{\mathcal{L}(F) : F \text{ is an arbitrary grammar form}\}.$
- (2) $\mathcal{M}_2 = \{\mathcal{L}(F) : F \text{ is a synchronised EOL form}\}.$
- (3) $\mathcal{M}_3 = \{\mathcal{L}(F) : F \text{ is a finite CF grammar form}\}.$
- (4) $\mathcal{M}_4 = \{\mathcal{L}(F) : F \text{ is a one sided linear grammar form}\}.$
- (5) $\mathcal{M}_5 = \{\mathcal{L} : \mathcal{L} \text{ is an s-grammatical family}\}.$

Proof. The proof of (1)–(5) is analogous to the one given for Example 1. In the proof of (5) it is of critical importance that condition (i) in the definition of an MSW space requires only closure under breaking but not closure under intersection with regular sets: by Ottmann *et al.* (1979), *s*-grammatical families are closed under breaking but not under intersection with regular sets.

EXAMPLE 3. Let \mathcal{G} be any collection of CF grammars such that

- (a) \mathcal{G} is closed under interpretation;
- (b) whenever F_1 and F_2 are in \mathcal{G} , then so is $F_1 \oplus F_2$.

Then $\mathcal{M}(\mathcal{G}) = \{\mathcal{L} : \mathcal{L} = \mathcal{L}(F), F \in \mathcal{G}\}$ is an MSW space.

Proof. (i) Since $\mathcal{L}(F)$ is closed under superdisjoint union and intersection by regular sets, condition (i) for MSW spaces is satisfied.

(ii) Because of assumption (b) above, condition (ii) for MSW spaces is satisfied.

(iii) For each infinite CF grammar F and $L = L(F)$, define $L_i = L(i) = \{x : x \in L, |x| \neq i\}$. Condition (iii) for MSW space is now established analogous to the approach in Example 1.

Example 3 leads in a natural way to the definition of a principal MSW space.

EXAMPLE 4. Let F be a grammar form and let $\mathcal{M} = \{\mathcal{L}(F') : F' \triangleleft F\}$. Then \mathcal{M} is an MSW space which we call a *principal* MSW space (*generated* by F).

Proof. \mathcal{M} is clearly interpretation closed. For any two $\mathcal{L}_1, \mathcal{L}_2$ of \mathcal{M} , $\mathcal{L}_1 \dot{\vee} \mathcal{L}_2$ is also in \mathcal{M} by using the obvious construction. Hence the situation is exactly the same as in Example 3.

EXAMPLE 5. $\mathcal{M} = \{\mathcal{L}_{\text{Reg}}\} \cup \{\mathcal{F}_t : t \geq 1\}$, where \mathcal{F}_t is the family of all finite languages none of which contains a word of length exceeding t , is an MSW space.

Proof. We consider the conditions in the definition of an MSW space one by one.

- (i) Each $\mathcal{L} \in \mathcal{M}$ is clearly closed under $\dot{\cup}$ and breaking.
- (ii) To see that \mathcal{M} is closed under $\dot{\vee}$ observe that $\mathcal{F}_t \dot{\vee} \mathcal{F}_s = \mathcal{F}_{\max(t,s)}$, that $\mathcal{F}_t \dot{\vee} \mathcal{L}_{\text{Reg}} = \mathcal{L}_{\text{Reg}}$ and that $\mathcal{L}_{\text{Reg}} \dot{\vee} \mathcal{L}_{\text{Reg}} = \mathcal{L}_{\text{Reg}}$.
- (iii) For each infinite language L define languages L_i by $L_i = \{x \in L : |x| \leq i\}$.

(a) If $L \in \mathcal{L} \in \mathcal{M}$, then clearly $\mathcal{L} = \mathcal{L}_{\text{Reg}}$ and hence $L_i \in \mathcal{L}$. If L occurs in some language family of \mathcal{M} , then L is regular. For $\mathcal{L} \in \mathcal{M}$, if $L_i \in \mathcal{L}$ for infinitely many i , then $\mathcal{L} = \mathcal{L}_{\text{Reg}}$. Thus $L \in \mathcal{L}$ as desired.

(b) For $\mathcal{L} \in \mathcal{M}$, if $L \in \mathcal{L}$ and for some p , $L_p \neq L$ then L contains words of length exceeding p , but L_p does not. Hence \mathcal{F}_p contains L_p but not L . ■

Concerning Example 5, note that $\mathcal{F}_t = \{L: L \in \mathcal{L}_{\text{reg}} \text{ and each word } x \text{ of } L \text{ satisfies } |x| \leq t\}$. Defining for an arbitrary language family \mathcal{L} the language families $\mathcal{L}_t (t \geq 1)$ by $\mathcal{L}_t = \{L': L' \in \mathcal{L} \text{ and each word } x \text{ of } L' \text{ satisfies } |x| \leq t\}$ we can clearly extend Example 5 as follows:

EXAMPLE 6. Let \mathcal{L} be an arbitrary family of languages which is closed under $\dot{\cup}$ and intersection by regular sets. Then $\mathcal{M} = \{\mathcal{L}\} \cup \{\mathcal{L}_t: t \geq 1\}$ is an MSW space.

Proof. Since \mathcal{M} contains a single language family \mathcal{L} which may contain infinite languages, the proof is exactly analogous to the one of Example 5.

After having established a number of auxiliary results we will later exhibit many more examples of MSW spaces. We will also show that a number of well-known collections of families are not MSW spaces.

We now turn to our first major theorem establishing that in an MSW space every "large enough" language family is density forcing.

THEOREM 3.1. Let \mathcal{M} be an MSW space and let \mathcal{F} be the collection of all finite languages occurring in language families of \mathcal{M} . If \mathcal{L} is a family of \mathcal{M} and \mathcal{L} contains \mathcal{F} , then \mathcal{L} is density forcing.

Proof. We have to show: if $\mathcal{L}_1, \mathcal{L}_2$ are arbitrary families of \mathcal{M} with $\mathcal{L} \subseteq \mathcal{L}_1 \subsetneq \mathcal{L}_2$, then there exists an $\mathcal{L}_3 \in \mathcal{M}$ such that $\mathcal{L}_1 \subsetneq \mathcal{L}_3 \subsetneq \mathcal{L}_2$ holds. Since we assume $\mathcal{L}_1 \subsetneq \mathcal{L}_2$, $\mathcal{L}_2 - \mathcal{L}_1$ contains a language L . Indeed, L must be infinite, since \mathcal{L} (and hence \mathcal{L}_1) contains all finite languages of \mathcal{M} . Furthermore, we may assume that L is coherent. (For if L is not coherent to start with, we have $L = L_1 \dot{\cup} L_2$ for some L_1 and L_2 , where $L_1 \neq \emptyset \neq L_2$. Since \mathcal{L}_1 is closed under $\dot{\cup}$ at least one of L_1 or L_2 is not in \mathcal{L}_1 . Assume without loss of generality that this is L_1 . But $L_1 \in \mathcal{L}_2$, since \mathcal{L}_2 is closed under breaking, i.e., $L_1 \in \mathcal{L}_2 - \mathcal{L}_1$ and must be infinite by the above argument. Repeating this process a finite number of times, an infinite, coherent L in $\mathcal{L}_2 - \mathcal{L}_1$ is finally obtained).

There exists a p with $L \neq L_p$ and $L_p \notin \mathcal{L}_1$. (Otherwise, if each $L_p \neq L$ is in \mathcal{L}_1 , L is in \mathcal{L}_1 by condition (iiia) for MSW spaces). But $L_p \in \mathcal{L}_2$ since $L \in \mathcal{L}_2$ and using condition (iiia) in the other direction. By condition (iiib) for MSW spaces there exists an $\mathcal{L}_p \in \mathcal{M}$ with $\mathcal{L}_p \subseteq \mathcal{L}_2$ such that \mathcal{L}_p contain L_p but does not contain L .

We are now ready to define the desired \mathcal{L}_3 by $\mathcal{L}_3 = \mathcal{L}_1 \dot{\vee} \mathcal{L}_p$. By condition (ii) for MSW spaces, $\mathcal{L}_3 \in \mathcal{M}$. Since $\mathcal{L}_3 = \mathcal{L}_1 \dot{\vee} \mathcal{L}_p$ is closed under breaking by condition (i) for MSW spaces, $\mathcal{L}_1 \subseteq \mathcal{L}_1 \dot{\vee} \mathcal{L}_p = \mathcal{L}_3$, and $L_p \in \mathcal{L}_1 \dot{\vee} \mathcal{L}_p$ for the same reason. But $L_p \notin \mathcal{L}_1$. Thus $\mathcal{L}_1 \subsetneq \mathcal{L}_3$. Since $\mathcal{L}_1 \subseteq \mathcal{L}_2$, $\mathcal{L}_p \subseteq \mathcal{L}_2$

and \mathcal{L}_2 is closed under superdisjoint union by condition (i) for MSW spaces, $\mathcal{L}_3 = \mathcal{L}_1 \dot{\vee} \mathcal{L}_p \subseteq \mathcal{L}_2$. Since $L \in \mathcal{L}_2$ but $L \notin \mathcal{L}_1$ and $L \notin \mathcal{L}_p$ and since L is coherent, $L \notin \mathcal{L}_1 \dot{\vee} \mathcal{L}_p$. Thus $\mathcal{L}_3 = \mathcal{L}_1 \dot{\vee} \mathcal{L}_p \subsetneq \mathcal{L}_2$, completing the proof. ■

Since we have discussed a variety of examples of MSW spaces satisfying the assumption of Theorem 3.1, the above theorem establishes the existence of a variety of collections of language families containing density forcing language families. Typically, any language family \mathcal{L} generated by a grammar form F such that \mathcal{L} contains all finite languages is density forcing, with respect to each $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$, where

$$\begin{aligned}\mathcal{M}_1 &= \{\mathcal{L}(F): F \text{ a CF grammar form}\}, \\ \mathcal{M}_2 &= \{\mathcal{L}(F): F \text{ a synchronized EOL form}\}, \\ \mathcal{M}_3 &= \{\mathcal{L}(F): F \text{ an arbitrary grammar form}\}.\end{aligned}$$

Theorem 3.1 is also an extremely useful tool for establishing that certain collections of language families are not MSW spaces:

EXAMPLE 7. $\mathcal{M}_G = \{\mathcal{L}: \mathcal{L} \text{ is generated by a grammar form in the sense of Cremers and Ginsburg}\}$ and

$$\mathcal{M}_{\text{AFL}} = \{\mathcal{L}: \mathcal{L} \text{ is an AFL}\}$$

are not MSW spaces.

Proof. For the definition of generation of language families by grammar forms in the Cremers–Ginsburg sense we refer to the pioneering paper Cremers and Ginsburg (1975). It is shown in that paper that \mathcal{M}_G contains \mathcal{L}_{Reg} and \mathcal{L}_{Lin} , but no language family inbetween. If \mathcal{M}_G were an MSW space, \mathcal{L}_{Reg} would be density forcing by Theorem 3.1, hence \mathcal{M}_G would contain (even infinitely many) language families between \mathcal{L}_{Reg} and \mathcal{L}_{Lin} , a contradiction. Similarly, if \mathcal{M}_{AFL} were an MSW space, every AFL \mathcal{L} would be density forcing (since \mathcal{L} contains all finite languages). But by, e.g., Theorem 6.6.2 of Ginsburg (1975) there exist AFLs \mathcal{L}_1 and \mathcal{L}_2 such that $\mathcal{L}_1 \subsetneq \mathcal{L}_2$, with no AFL between \mathcal{L}_1 and \mathcal{L}_2 . ■

The next theorem allows us to show that collections of language families derived from forms are by no means the only collections forming MSW spaces. Rather, “most” MSW spaces which we can obtain are obtained quite independently of form theory.

THEOREM 3.2. *Let \mathcal{M} be a collection of language families such that each family \mathcal{L} of \mathcal{M} is closed under superdisjoint union, intersection with regular sets and covering. Let $\bar{\mathcal{M}}$ be the closure of \mathcal{M} under superdisjoint wedge and extraction.*

Then each $\mathcal{L} \in \bar{\mathcal{M}}$ is closed under superdisjoint union, intersection with regular sets and covering, and $\bar{\mathcal{M}}$ is an MSW space.

Proof. We first show that each $\mathcal{L} \in \bar{\mathcal{M}}$ has the specified closure properties. It clearly suffices to show that superdisjoint wedge and extraction preserve the desired closure properties.

Suppose \mathcal{L}_1 and \mathcal{L}_2 are elements of $\bar{\mathcal{M}}$ and have the desired closure properties. We then show that $\mathcal{L} = \mathcal{L}_1 \dot{\vee} \mathcal{L}_2$ also has the described closure properties. Recall that $\mathcal{L} = \mathcal{L}_1 \dot{\vee} \mathcal{L}_2 = \{L_1 \dot{\cup} L_2 : L_1 \in \mathcal{L}_1, L_2 \in \mathcal{L}_2\}$ and that $L(i) = \{x \in L : |x| \neq i\}$ for all $i \geq 1$ and every language L .

Closure under $\dot{\cup}$

Consider arbitrary $Z_1, Z_2 \in \mathcal{L}$. Then $Z_1 = L_1 \dot{\cup} L_2, Z_2 = L'_1 \dot{\cup} L'_2$ with $L_1, L'_1 \in \mathcal{L}_1$ and $L_2, L'_2 \in \mathcal{L}_2$. Consider now $Z_1 \dot{\cup} Z_2 = (L_1 \dot{\cup} L_2) \dot{\cup} (L'_1 \dot{\cup} L'_2) = (L_1 \dot{\cup} L'_1) \dot{\cup} (L_2 \dot{\cup} L'_2)$. Note that the four alphabets involved are all disjoint. Since $L_1, L'_1 \in \mathcal{L}_1$, $Y_1 = L_1 \dot{\cup} L'_1$ is in \mathcal{L}_1 and similarly $Y_2 = L_2 \dot{\cup} L'_2 \in \mathcal{L}_2$. Thus, $Z_1 \dot{\cup} Z_2 = Y_1 \dot{\cup} Y_2$ with $Y_1 \in \mathcal{L}_1, Y_2 \in \mathcal{L}_2$, i.e., $Z_1 \dot{\cup} Z_2 \in \mathcal{L}$.

Closure under Intersection with Regular Sets

Consider an arbitrary L in \mathcal{L} , i.e., $L = L_1 \dot{\cup} L_2$ with $L_1 \in \mathcal{L}_1, L_2 \in \mathcal{L}_2$, and consider an arbitrary regular set R . We want to show that $L \cap R \in \mathcal{L}$. But $L \cap R = (L_1 \cap R) \dot{\cup} (L_2 \cap R)$. Since $\mathcal{L}_1, \mathcal{L}_2$ are closed under intersection with regular sets, $L \cap R = L'_1 \dot{\cup} L'_2$ with $L'_1 \in \mathcal{L}_1, L'_2 \in \mathcal{L}_2$, i.e., $L \cap R \in \mathcal{L}$, as desired.

Closure under Covering

Consider an arbitrary L such that $L(i)$ is in \mathcal{L} for infinitely many i . We have to show $L \in \mathcal{L}$. For each i we have $L(i) = L_1^{(i)} \dot{\cup} L_2^{(i)}$ with $L_1^{(i)} \in \mathcal{L}_1, L_2^{(i)} \in \mathcal{L}_2$. Observe that L is over some fixed finite alphabet Σ which allows only finitely many different partitions $\Sigma = \Sigma_1 \dot{\cup} \Sigma_2$. Hence there is one partition, say $\Sigma = \Sigma_1 \dot{\cup} \Sigma_2$, such that for infinitely many i we have: $L(i) = L_1^{(i)} \dot{\cup} L_2^{(i)}$, where $L_1^{(i)} \subseteq \Sigma_1^*, L_2^{(i)} \subseteq \Sigma_2^*$, i.e., where $L_1^{(i)} = M_i^{(1)}(i)$ with $M^{(1)} = L \cap \Sigma_1^*$ and $L_2^{(i)} = M^{(2)}(i)$ with $M^{(2)} = L \cap \Sigma_2^*$. Since for infinitely many i , $M^{(1)}(i) \in \mathcal{L}_1$ and $M^{(2)}(i) \in \mathcal{L}_2$ we have $M^{(1)} \in \mathcal{L}_1$ and $M^{(2)} \in \mathcal{L}_2$, since \mathcal{L}_1 and \mathcal{L}_2 are closed under covering. But $L = M^{(1)} \dot{\cup} M^{(2)}$, hence $L \in \mathcal{L}$ as claimed.

Consider now an $\mathcal{L} \in \bar{\mathcal{M}}$ with the desired closure properties. We have to show that $\mathcal{L}(p) = \{L(p) : L \in \mathcal{L}\}$ also has the described closure properties.

Closure under $\dot{\cup}$

Consider two languages $L^{(1)}, L^{(2)}$ in $\mathcal{L}(p)$. We have to show $L^{(1)} \dot{\cup} L^{(2)}$ is again in $\mathcal{L}(p)$. We have $L^{(1)} = M^{(1)}(p)$ and $L^{(2)} = M^{(2)}(p)$ for some languages

$M^{(1)}, M^{(2)}$ in \mathcal{L} . Since \mathcal{L} is closed under intersection with regular languages, $M^{(1)}(p) \dot{\cup} M^{(2)}(p) = L^{(1)} \dot{\cup} L^{(2)}$ is in \mathcal{L} , and clearly in $\mathcal{L}(p)$.

Closure under Intersection with Regular Sets

Consider a language L in $\mathcal{L}(p)$ and a regular set R . We have to show that $L \cap R$ is again in $\mathcal{L}(p)$. Now $L = M(p)$ for some $M \in \mathcal{L}$, hence $M \cap R \in \mathcal{L}$, hence $(M \cap R)(p) \in \mathcal{L}$, i.e., $L \cap R \in \mathcal{L}$ and indeed in $\mathcal{L}(p)$.

Closure under Covering

Suppose \mathcal{L} is closed under covering. We have to show that $\mathcal{L}(p)$ is also closed under covering. Take an arbitrary language L such that $L(i)$ is in $\mathcal{L}(p)$ for infinitely many i . (Hence L cannot contain words of length p). Since $\mathcal{L}(p) \subseteq \mathcal{L}$, all these $L(i)$ are in \mathcal{L} , hence L is in \mathcal{L} . Hence $L(p) = L \in \mathcal{L}(p)$, as desired.

To complete the proof it remains to show that $\bar{\mathcal{M}}$ is an MSW space. We show that all conditions of the definition of an MSW space are indeed met.

(i) That $\mathcal{L} \in \bar{\mathcal{M}}$ is closed under $\dot{\cup}$ and breaking is clear, since breaking is a special case of intersection with regular sets.

(ii) $\bar{\mathcal{M}}$ is closed under $\dot{\vee}$ by construction.

(iii) For each infinite language L we define once more $L_i = L(i) = \{x \in L: |x| \neq i\}$.

(a) Suppose $L(i) \neq L$ implies $L(i) \in \mathcal{L}$. Since \mathcal{L} is closed under covering, $L \in \mathcal{L}$. Suppose $L \in \mathcal{L}$, then $L(i) \in \mathcal{L}$ for all i since \mathcal{L} is closed under intersection with regular sets.

(b) Consider an arbitrary $L \in \bar{\mathcal{M}}$ and some $L(p) \neq L$. The language family $\mathcal{L}(p) = \{L'(p): L' \in \mathcal{L}\}$ is in $\bar{\mathcal{M}}$ by construction of $\bar{\mathcal{M}}$, $\mathcal{L}(p) \subseteq \mathcal{L}$, $L(p) \in \mathcal{L}(p)$ but $L \notin \mathcal{L}(p)$.

COROLLARY 3.3. *Consider an arbitrary (finite or infinite) collection \mathcal{S} of languages. Close each language $L \in \mathcal{S}$ with respect to the operations $\dot{\cup}$, \cap with regular sets, and covering, yielding a collection \mathcal{M} of language families. Close \mathcal{M} under superdisjoint wedge and extraction to obtain $\bar{\mathcal{M}}$. Then $\bar{\mathcal{M}}$ is an MSW space.*

Proof. Clear.

COROLLARY 3.4. *Let $\bar{\mathcal{M}}$ be the collection of all language families consisting of CF languages such that each $\mathcal{L} \in \bar{\mathcal{M}}$ is closed under $\dot{\cup}$, intersection with regular sets, and covering. Then $\bar{\mathcal{M}}$ is an MSW space.*

Proof. The closure of $\bar{\mathcal{M}}$ under $\dot{\vee}$ and extraction is again $\bar{\mathcal{M}}$.

COROLLARY 3.5. *Let \mathcal{S} be an arbitrary family of languages closed under $\dot{\cup}$*

and \cap with regular sets. Let $\bar{\mathcal{M}}$ be the collection of all those subsets \mathcal{L} of \mathcal{S} which are closed under \cup , \cap with regular sets and covering. Then $\bar{\mathcal{M}}$ is an MSW space.

Proof. It suffices to show that $\bar{\mathcal{M}}$ is closed under $\dot{\vee}$ and extraction. First consider two families $\mathcal{L}_1, \mathcal{L}_2$ in $\bar{\mathcal{M}}$. Since $\mathcal{L}_1 \dot{\vee} \mathcal{L}_2 = \{L_1 \dot{\cup} L_2 : L_1 \in \mathcal{L}_1, L_2 \in \mathcal{L}_2\}$ and since $\mathcal{L}_1 \subseteq \mathcal{S}, \mathcal{L}_2 \subseteq \mathcal{S}$ and \mathcal{S} is closed under union $\mathcal{L}_1 \dot{\vee} \mathcal{L}_2 \subseteq \mathcal{S}$. Since $\mathcal{L}_1 \dot{\vee} \mathcal{L}_2$ is closed under $\dot{\cup}, \cap$ with regular sets and covering (as shown in the proof of Theorem 3.2), $\mathcal{L}_1 \dot{\vee} \mathcal{L}_2 \in \bar{\mathcal{M}}$ since $\bar{\mathcal{M}}$ contains all such sets. Similarly, if $\mathcal{L} \in \bar{\mathcal{M}}$ then $\mathcal{L}(p) \in \bar{\mathcal{M}}$, since $\mathcal{L}(p)$ is closed under \cup, \cap with regular sets and covering.

Our next aim is to show that closing a language family under the operations intersection with regular sets, covering and superdisjoint union can be carried out by using the operations only in that order. Similarly we show that closing a collection of language families under the operations of extraction and superdisjoint wedge can be carried out by first using only extractions, then only wedges. We then combine these facts with Theorem 3.2 into a simple tool (Theorem 3.10) for constructing MSW spaces.

We start by establishing the mentioned auxiliary results.

LEMMA 3.6. *Let \mathcal{L} be an arbitrary language family and let $\bar{\mathcal{L}}$ be its closure under $\dot{\cup}$, i.e., $\bar{\mathcal{L}} = \{L : L = L_1 \dot{\cup} L_2 \dot{\cup} \dots \dot{\cup} L_n, n \geq 1, L_i \in \mathcal{L} \text{ for } 1 \leq i \leq n\}$. If \mathcal{L} is closed under intersection by regular sets and covering, then so is $\bar{\mathcal{L}}$.*

Proof. Suppose $L \in \bar{\mathcal{L}}$ and suppose R is a regular set. Then $L = L_1 \dot{\cup} L_2 \dot{\cup} \dots \dot{\cup} L_n$ for some $n \geq 1$ and $L_i \in \mathcal{L}$. We have

$$L \cap R = (L_1 \cap R) \dot{\cup} (L_2 \cap R) \dot{\cup} \dots \dot{\cup} (L_n \cap R).$$

Since $L_i \cap R = L'_i$ is in \mathcal{L} by the assumption for $i = 1, 2, \dots, n$ $L \cap R \in \bar{\mathcal{L}}$, as required.

Suppose for some $L, L \subseteq \Sigma^*$, $L(i)$ is in $\bar{\mathcal{L}}$ for infinitely many i . Since Σ has only a finite number of partitions $\Sigma_1 \dot{\cup} \Sigma_2 \dot{\cup} \dots \dot{\cup} \Sigma_k$ with $k \geq 1$ and $\Sigma_i \neq \emptyset$ for $i = 1, 2, \dots, k$ there exists one partition, say $\Sigma_1 \dot{\cup} \Sigma_2 \dot{\cup} \dots \dot{\cup} \Sigma_k$ such that $L(i) = L_1^{(i)} \dot{\cup} L_2^{(i)} \dot{\cup} \dots \dot{\cup} L_k^{(i)}$ with $L_t^{(i)} \subseteq \Sigma_t^*$ ($1 \leq t \leq k$) holds for infinitely many i , where $L_t^{(i)} = M_t^{(i)}(i)$, $L_t^{(i)} \in \mathcal{L}$ for each $t = 1, 2, \dots, k$ and where $M^{(i)} = L \cap \Sigma_i^*$. Since \mathcal{L} is closed under covering, $M^{(i)}$ in \mathcal{L} for $t = 1, 2, \dots, k$. Since $L = M^{(1)} \dot{\cup} M^{(2)} \dot{\cup} \dots \dot{\cup} M^{(k)}$, $L \in \bar{\mathcal{L}}$, as required. ■

LEMMA 3.7. *Let \mathcal{L} be a language family and $\bar{\mathcal{L}}$ its closure under covering. If \mathcal{L} is closed under intersection with regular sets then so is $\bar{\mathcal{L}}$.*

Proof. Suppose an L in $\bar{\mathcal{L}}$ and a regular set R are chosen arbitrarily. We have to show that $L \cap R$ is in $\bar{\mathcal{L}}$.

We distinguish two cases:

(a) L is in \mathcal{L} . Then $L \cap R$ is in \mathcal{L} and hence $L \cap R$ is in \mathcal{P} .

(b) Suppose L is not in \mathcal{L} . Since L is in \mathcal{P} , $L(i)$ is in \mathcal{L} for infinitely many i . Hence $L(i) \cap R(i) \in \mathcal{L}$ holds for all those i . But $L(i) \cap R(i) = (L \cap R)(i)$, i.e., $(L \cap R)(i) \in \mathcal{L}$ for infinitely many i . This implies $L \cap R \in \mathcal{P}$, as required. ■

COROLLARY 3.8. *To close a language family under intersection by regular sets, covering and superdisjoint union it suffices to first close it under intersection by regular sets, then close it under covering and then close it under superdisjoint union.*

LEMMA 3.9. *Let \mathcal{M} be a language family and let $\bar{\mathcal{M}}$ be its closure under superdisjoint wedge. If \mathcal{M} is closed under extraction, then so is $\bar{\mathcal{M}}$.*

Proof. Let \mathcal{P} be an arbitrary language family of $\bar{\mathcal{M}}$, i.e.,

$$\mathcal{P} = \{L_1 \dot{\cup} L_2 \dot{\cup} \cdots \dot{\cup} L_k : L_t \in \mathcal{L}_t, \mathcal{L}_t \in \mathcal{M}, t \geq 1\}.$$

We have to show that $\mathcal{P}(i) = \{L(i) : L \in \mathcal{P}\}$ is also in $\bar{\mathcal{M}}$. Since \mathcal{M} is closed under extraction,

$$\mathcal{P}(i) = \{L_1(i) \dot{\cup} L_2(i) \dot{\cup} \cdots \dot{\cup} L_k^{(i)} : L_t \in \mathcal{L}_t, \mathcal{L}_t \in \mathcal{M}, t \geq 1\}$$

implies

$$\begin{aligned} \mathcal{P}(i) &= \{L'_1 \dot{\cup} L'_2 \dot{\cup} \cdots \dot{\cup} L'_k : L'_t \in \mathcal{L}_t(i), \mathcal{L}_t(i) \in \mathcal{M}, t \geq 1\} \\ &= \mathcal{L}_1(i) \dot{\vee} \mathcal{L}_2(i) \dot{\vee} \cdots \dot{\vee} \mathcal{L}_k(i), \end{aligned}$$

i.e., $\mathcal{P}(i) \in \bar{\mathcal{M}}$, as desired. ■

Corollary 3.8, Lemma 3.9 and Theorem 3.2 are now combined into Theorem 3.10.

THEOREM 3.10. *Let \mathcal{M} be an arbitrary collection of families of languages. Close each family \mathcal{L} of \mathcal{M} first under intersection by regular sets, then under covering, then under superdisjoint union. Close the resulting collection $\bar{\mathcal{M}}$ of language families under extraction and then under superdisjoint wedge. The resulting collection $\bar{\bar{\mathcal{M}}}$ of language families is an MSW space.*

Proof. Clear.

EXAMPLE 8. Let $\mathcal{M} = \{\mathcal{L}\}$, $\mathcal{L} = \{L\}$, $L = \Sigma^*$ for some alphabet Σ . We obtain $\bar{\mathcal{M}} = \{\mathcal{L}\}$, where \mathcal{L} contains all languages of the form $R_1 \dot{\cup} R_2 \dot{\cup} \cdots \dot{\cup} R_k$, where R_t ($t = 1, 2, \dots, k$) are regular sets over an alphabet of size $|\Sigma|$.

Closing \mathcal{M} under extraction yields a collection of language families \mathcal{M}' , where

$$\mathcal{M}' = \{\mathcal{L}_{i_1, i_2, \dots, i_m} : 1 \leq i_1 < i_2 < \dots < i_m, m \geq 0\},$$

where

$$\mathcal{L}_{i_1, i_2, \dots, i_m} = \{(\dots(L(i_1))(i_2)) \dots)(i_m) : L \in \mathcal{L}\}.$$

Closing $\bar{\mathcal{M}}$ under superdisjoint wedge gives an MSW space by Theorem 3.10.

Observe that many regular languages do not occur in any language family of $\bar{\mathcal{M}}$: if L is coherent and over an alphabet of more than k symbols, then L is in no \mathcal{L} of $\bar{\mathcal{M}}$.

RECEIVED: February 25, 1980; REVISED: September 4, 1980

REFERENCES

- CREMERS, A. B. AND GINSBURG, S. (1975), Context-free grammar forms, *J. Comput. Systems Sci.* **11**, 86–116.
- GINSBURG, S. (1975), "Algebraic and Automata-Theoretic Properties of Eormal Languages," North-Holland, Amsterdam.
- HARRISON, M. A. (1978), "Introduction to Formal Language Theory," Addison-Wesley, Reading, Mass.
- MAURER, H. A., SALOMAA, A., AND WOOD, D. (1977), EOL forms, *Acta Inform.* **8**, 75–96.
- MAURER, H. A., SALOMAA, A., AND WOOD, D. (1980a), Context-Free Grammar Forms with Strict Interpretations, *J. Comput. Systems Sci.* **21**, 110–135.
- MAURER, H. A., SALOMAA, A., AND WOOD, D. (1980b), Synchronized EOL Forms, *Theoret. Comput. Sci.* **12**, 135–159.
- MAURER, H. A., SALOMAA, A., WOOD, D. (1979b), "Uniform Interpretations of Grammar Forms," Report 28, Institut für Informationsverarbeitung, Technische Universität Graz.
- MAURER, H. A., SALOMAA, A., AND WOOD, D. (1979c), "Dense Hierarchies of Grammatical Families," Report 30, Institut für Informationsverarbeitung, Technische Universität Graz; *Assoc. Comput. Mach.*, in press.
- MAURER, H. A., SALOMAA, A., AND WOOD, D. (1979d), "Decidability and Density in Two-Symbol Grammar Forms," Report 38, Institut für Informationsverarbeitung, Technische Universität Graz.
- MAURER, H. A., PENTONNEN, M., SALOMAA, A., AND WOOD, D. (1979), On noncontext-free grammar forms, *Math. Systems Theory*, 297–324.
- MAURER, H. A. (1969), "Theoretische Grundlagen der Programmiersprachen," BI, Mannheim.
- OTTMANN, TH., SALOMAA, A., WOOD, D. (1979), "Grammar and s-Grammar Forms: Decidability and Density," McMaster University Computer Science Report.
- SALOMAA, A. (1975), "Formal Languages," Academic Press, New York.
- WOOD, D. (1979), A survey of grammar and L forms, in "Proceedings, MFCS' 79," Lecture Notes in Computer Science No. 74, pp. 191–200, Springer-Verlag, Berlin/New York.
- WOOD, D. (19780), "Grammar and L forms: An Introduction," Lecture Notes in Computer Science No. 91, Springer-Verlag, Berlin/New York.